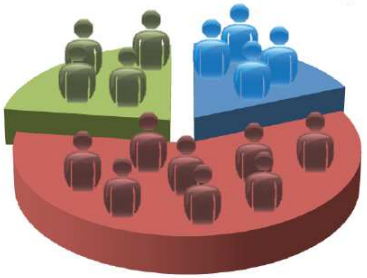
Labor_11

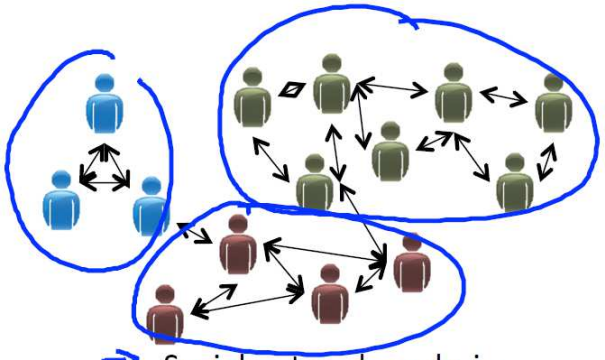
K-MEANS ALGORITHM

Unsupervised learning

Applications of clustering

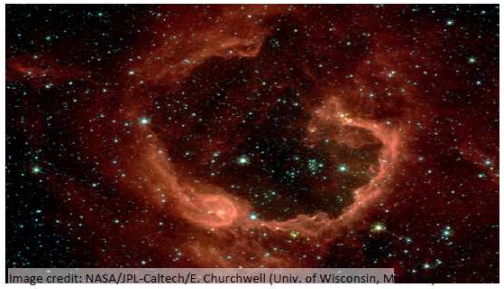


→ Market segmentation



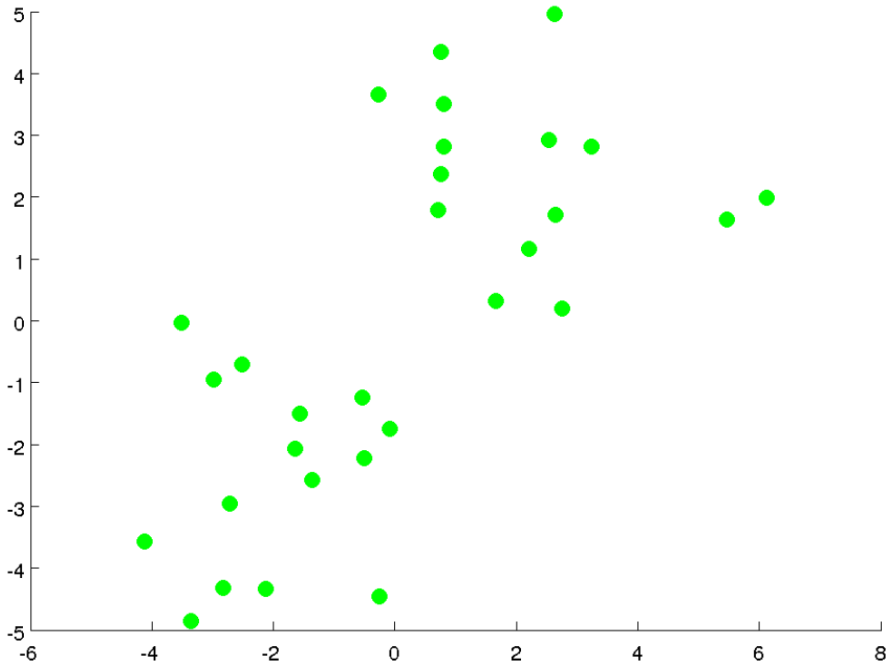
→ Social network analysis

→ Organize computing clusters

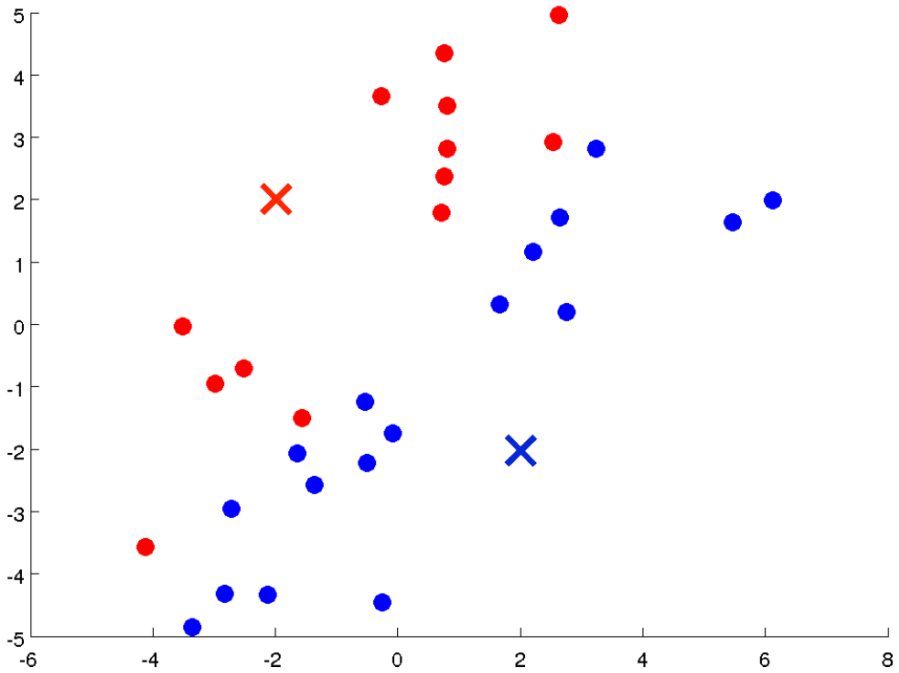


→ Astronomical data analysis

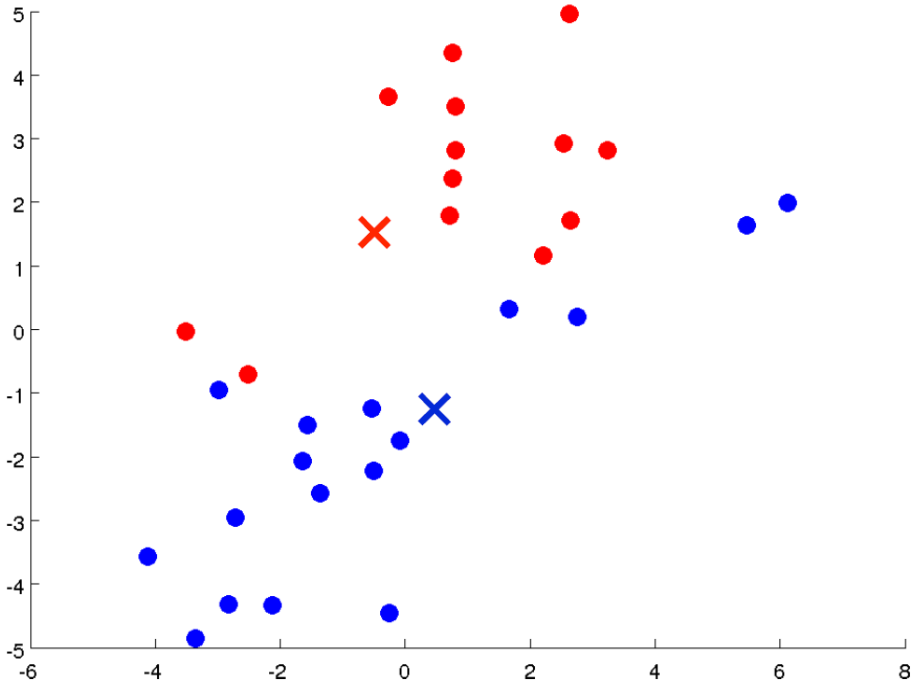
K-Means Algorithm example



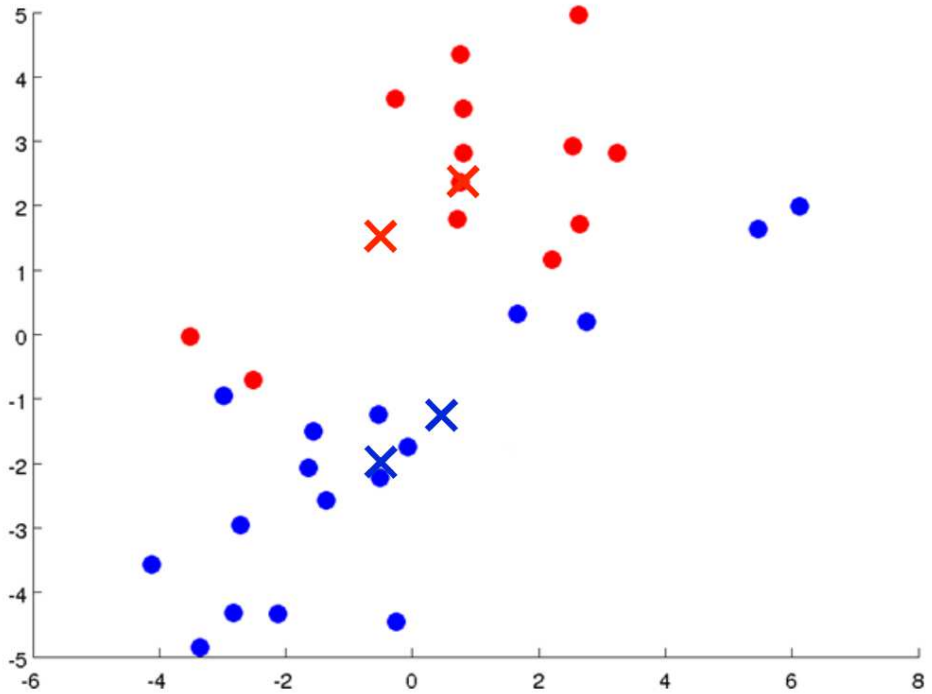
K-Means Algorithm example



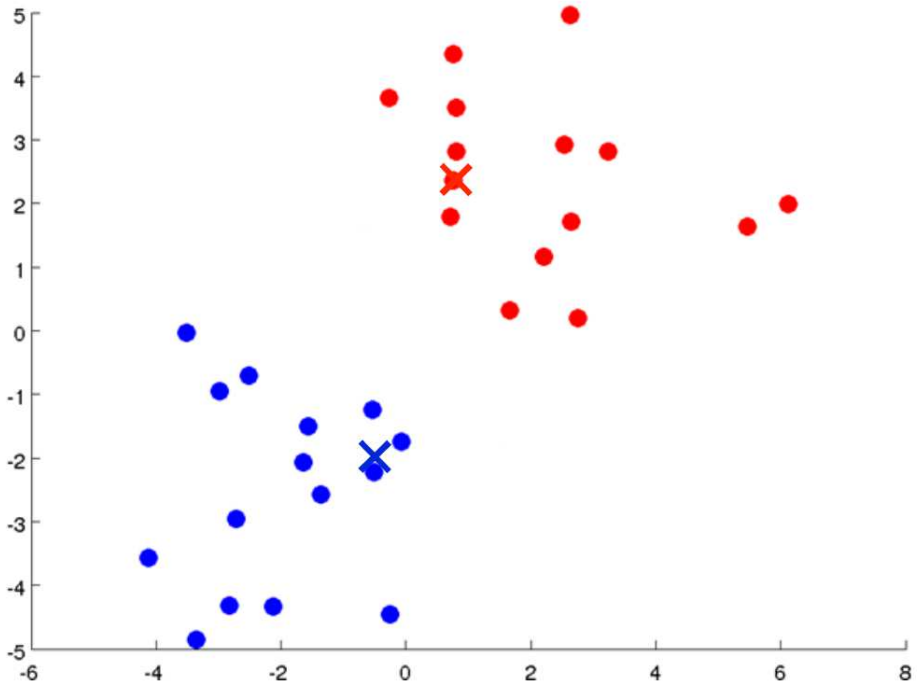
K-Means Algorithm example



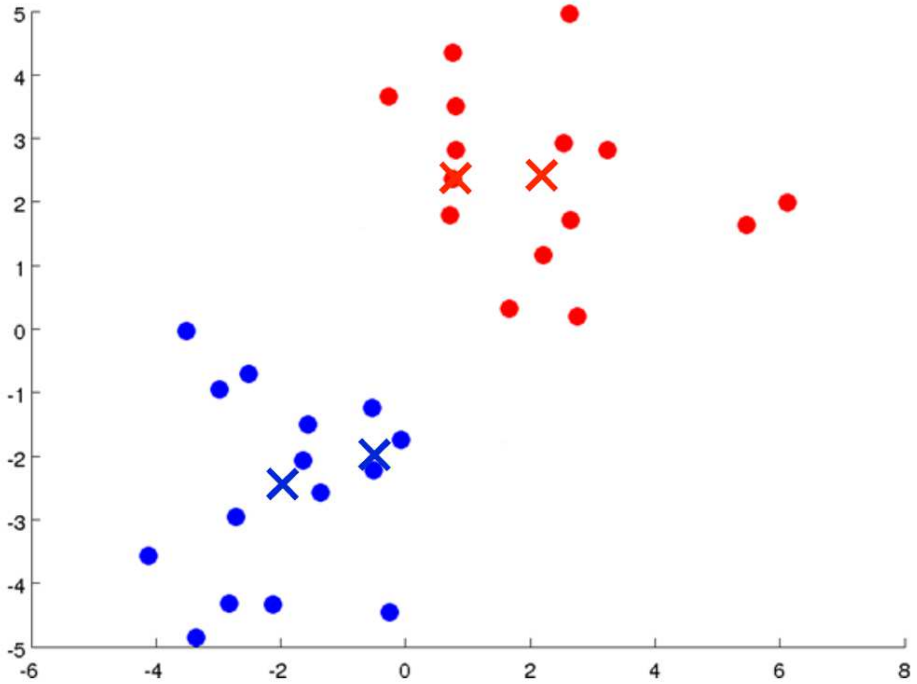
K-Means Algorithm example



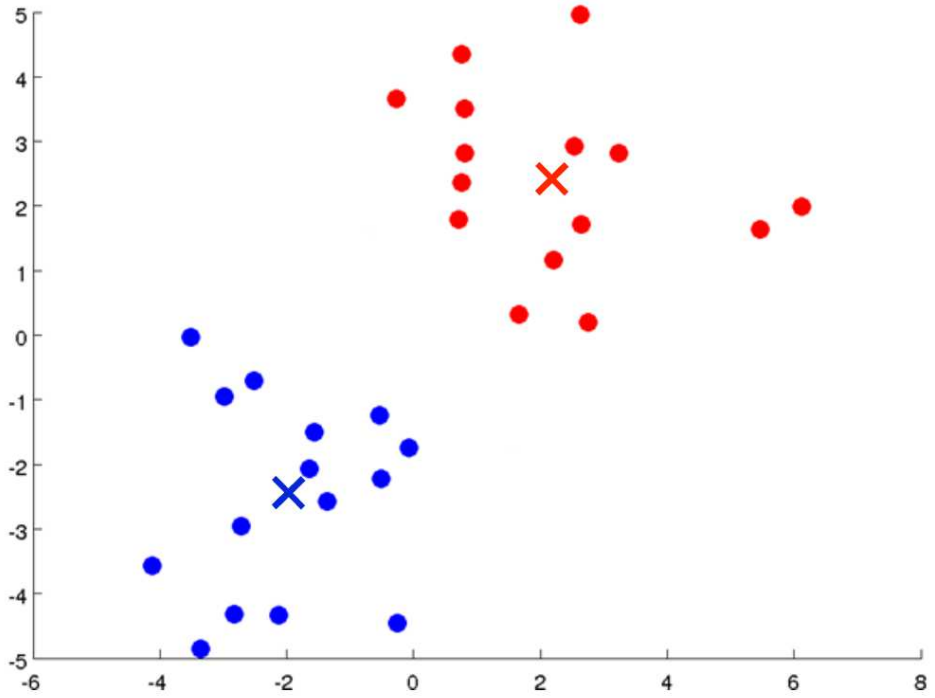
K-Means Algorithm example



K-Means Algorithm example



K-Means Algorithm example



K-Means pseudocode

Randomly initialize K cluster centroids $\mu_1, \mu_2, \ldots, \mu_K \in \mathbb{R}^n$

Repeat {

for i = 1 to m

$c^{(i)}$:= index (from 1 to K) of cluster centroid closest to $x^{(i)}$

for k = 1 to K

μ_k := average (mean) of points assigned to cluster k

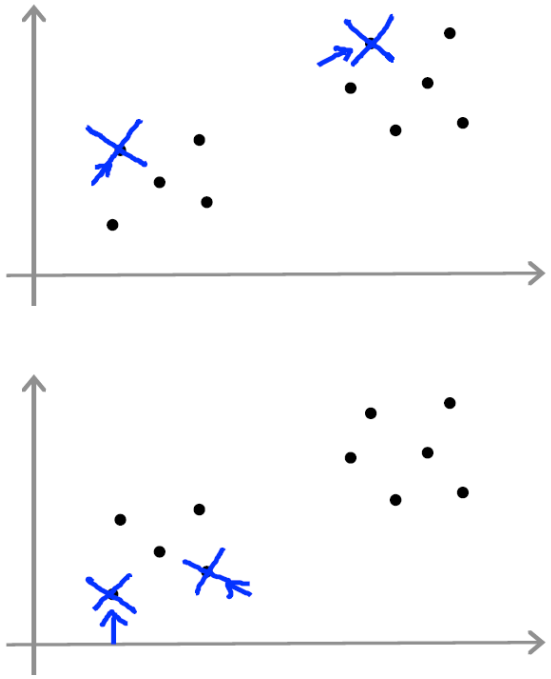
}

Random initialization

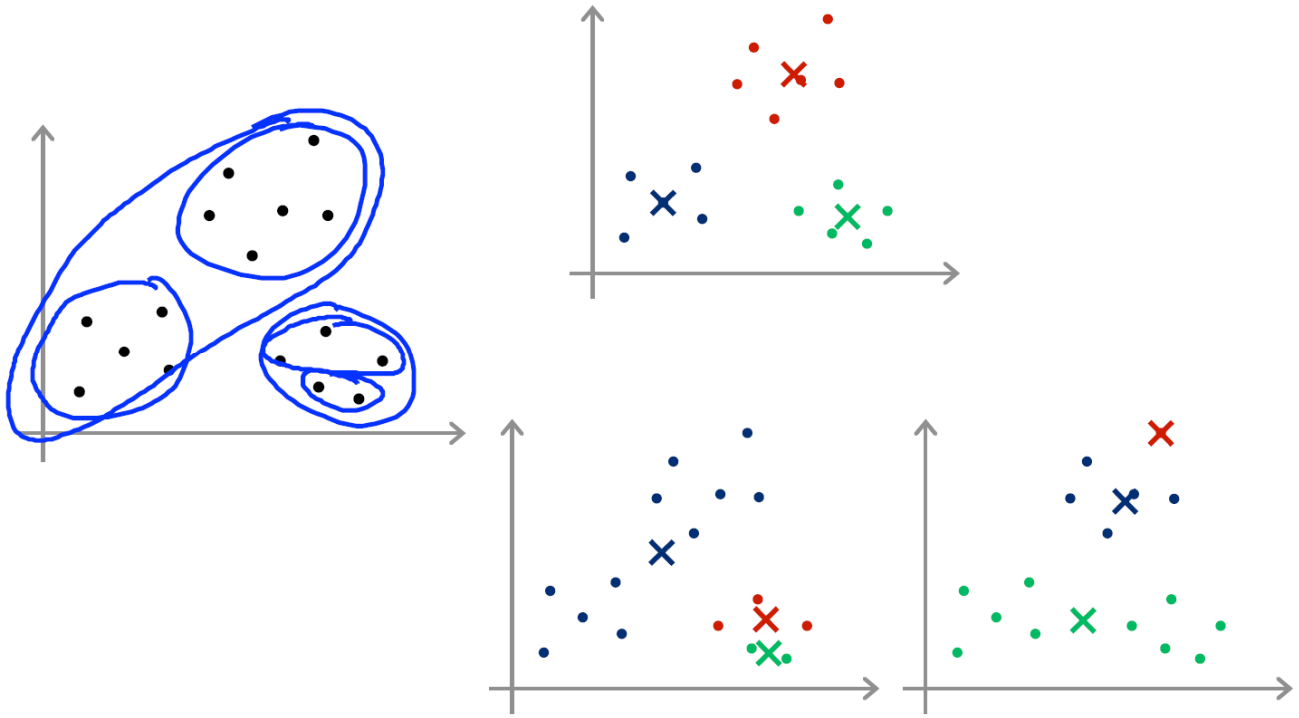
Should have $K < m$

Randomly pick K training examples.

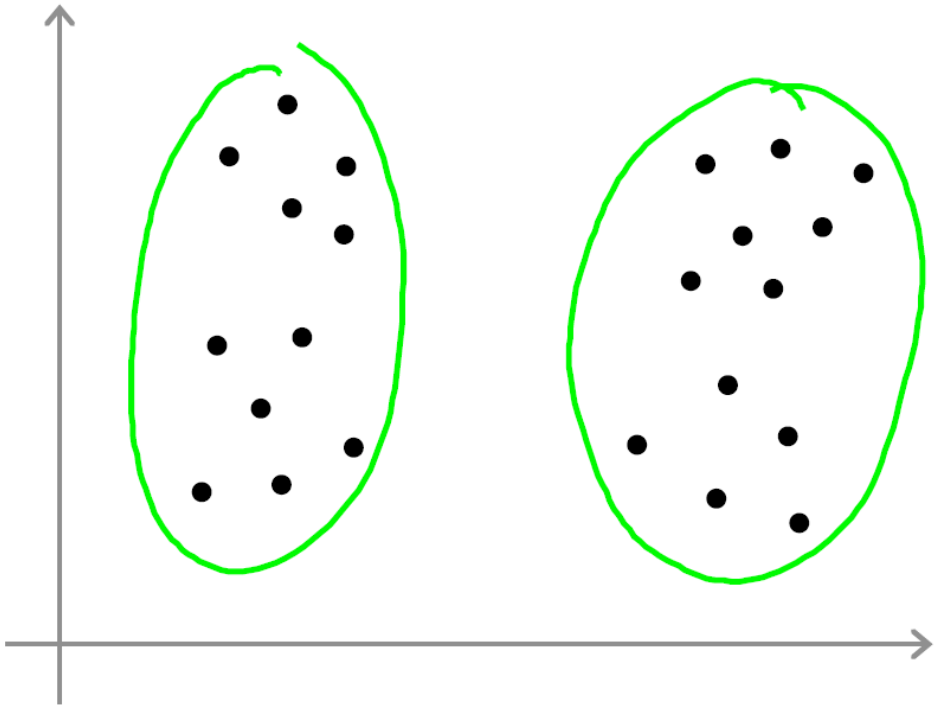
Set $\mu_1, \ldots, \mu_K$ equal to these K examples.



Local optimum

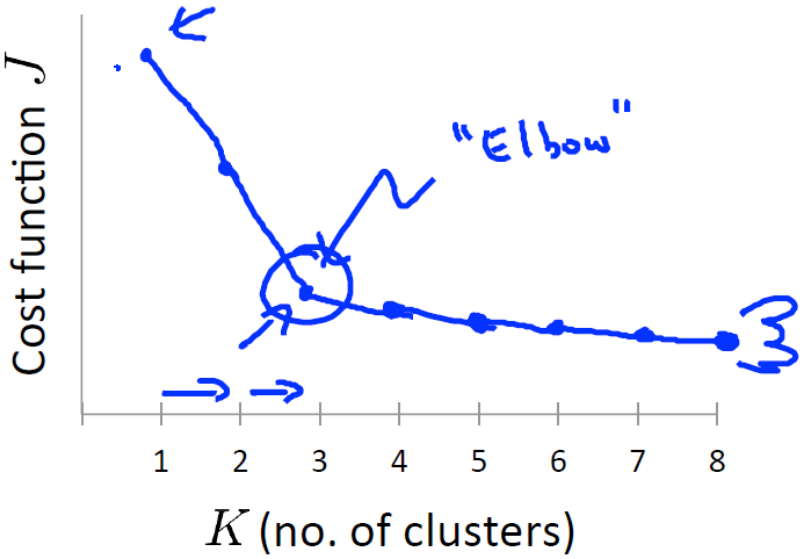


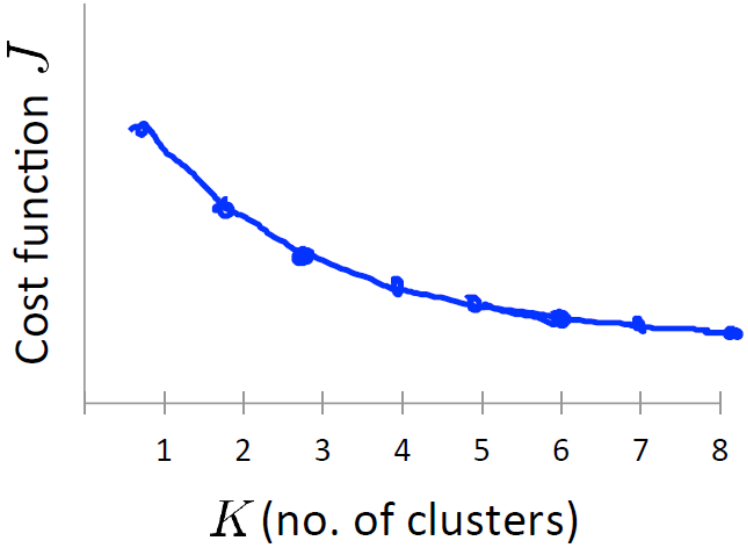
What is the right number of K?



Choosing K

Elbow method:





Choosing K

Sometimes, you're running K-means to get clusters to use for some later/downstream purpose. Evaluate K-means based on a metric for how well it performs for that later purpose.

E.g.